Biol Blood Marrow Transplant 21 (2015) 761–767

Biology of Blood and Marrow Transplantation

journal homepage: www.bbmt.org

ASBMT American Society for Blood and Marrow Transplantation

A Refined Risk Score for Acute Graft-versus-Host Disease that Predicts Response to Initial Therapy, Survival, and Transplant-Related Mortality

Margaret L. MacMillan [1,2,*], Marie Robin [3], Andrew C. Harris [4], Todd E. DeFor [1,5], Paul J. Martin [6], Amin Alousi [7,8], Vincent T. Ho [8,9], Javier Bolaños-Meade [8,10], James L.M. Ferrara [4,8], Richard Jones [8,10], Mukta Arora [1,11], Bruce R. Blazar [1,2], Shernan G. Holtan [1,11], David Jacobsohn [8,12], Marcelo Pasquini [8,13], Gerard Socie [3], Joseph H. Antin [8,9], John E. Levine [4,8], Daniel J. Weisdorf [1,8,11]

[1] *Blood and Marrow Transplant Program, University of Minnesota, Minneapolis, Minnesota*
[2] *Department of Pediatrics, University of Minnesota, Minneapolis, Minnesota*
[3] *Hôpital Saint Louis, Paris, France*
[4] *Blood and Marrow Transplant Program, University of Michigan, Ann Arbor, Michigan*
[5] *Department of Biostatistics and Informatics Core, Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota*
[6] *Fred Hutchinson Cancer Research Center, University of Washington, Seattle, Washington*
[7] *MD Anderson Cancer Center, Houston, Texas*
[8] *Blood and Marrow Transplant Clinical Trials Network GVHD Study Committee, Milwaukee, Wisconsin*
[9] *Dana-Farber Cancer Institute, Boston, Massachusetts*
[10] *Johns Hopkins University, Baltimore, Maryland*
[11] *Department of Medicine, University of Minnesota, Minneapolis, Minnesota*
[12] *Children's National Medical Center, Washington, District of Columbia*
[13] *Medical College of Wisconsin, Milwaukee, Wisconsin*

Article history:
Received 11 November 2014
Accepted 2 January 2015

Key Words:
Graft-versus-host disease
Grading systems
Risk score
Transplant-related mortality
Survival
Allogeneic hematopoietic cell transplantation

A B S T R A C T

To develop a novel acute graft-versus-host disease (GVHD) risk score, we examined the GVHD clinical stage and grade of 1723 patients at the onset of treatment with systemic steroids. Using clinical grouping, descriptive statistics and recursive partitioning, we identified poorly responsive, high-risk (HR) acute GVHD by the number of involved organs and severity of GVHD at onset. The overall response (complete response/partial response) rate 28 days after initiation of steroid therapy for acute GVHD was lower in the 269 patients with HR-GVHD than in the 1454 patients with standard risk (SR)-GVHD (44% [95% confidence interval (CI) 38% to 50%] versus 68% [95% CI, 66% to 70%], $P < .001$). Patients with HR-GVHD were less likely to respond at day 28 (odds ratio [OR], .3; 95% CI, .2 to .4; $P < .001$) and had higher risks of mortality (relative risk, 2.1; 95% CI, 1.7 to 2.6; $P < .001$) and transplant-related mortality (relative risk, 2.5; 95% CI, 2.0% to 3.2%, $P < .001$) than patients with SR-GVHD. This refined definition of acute GVHD risk is a better predictor of response, survival, and transplant-related mortality than other published acute GVHD risk scores. Patients with HR-GVHD are candidates for studies investigating new treatment approaches. Likewise, patients with SR-GVHD are candidates for studies investigating less toxic therapy.

INTRODUCTION

Acute graft-versus-host disease (GVHD) remains a major cause of morbidity and mortality after hematopoietic cell transplantation (HCT). Corticosteroids are the standard initial therapy but are effective in only approximately half of the cases [1,2]. Methods to identify patients who are unlikely to respond to this conventional initial therapy and who warrant alternative, more effective initial therapy are needed. Earlier analyses have defined high-risk (HR) acute GVHD at time of onset by clinical stage and grade [2-5] and, more recently, by serum biomarkers [6-8].

The Minnesota group recently defined HR acute GVHD by a novel acute GVHD risk score [9]. Initial high-risk (HR) acute GVHD was defined as either skin stage 4; lower

Financial disclosure: See Acknowledgments on page 767.

* Correspondence and reprint requests: Margaret L. MacMillan, MD, Department of Pediatrics, University of Minnesota, MMC 484, 420 Delaware Street, S.E., Minneapolis, MN 55455.
E-mail address: macmi002@umn.edu (M.L. MacMillan).

http://dx.doi.org/10.1016/j.bbmt.2015.01.001
1083-8791/

gastrointestinal (GI) stage 3-4 or liver stage 3-4; or skin stage 3+ and either lower GI 2-4 or liver stage 2-4 GVHD. Patients with this HR-GVHD were less likely to respond to steroid therapy and had a 2-fold increased risk of treatment-related (non-relapse) mortality (TRM) compared with patients with standard-risk(SR)-GVHD. In order to validate this new GVHD risk score, we examined a larger, heterogeneous group of patients who received steroids as initial systemic therapy for acute GVHD. A database of 1723 patients was created from 5 cohorts: the previously reported patients from the University of Minnesota (n = 864) [9], Blood and Marrow Transplant Clinical Trials Network (BMT CTN) study 0302 (n = 155) [10], BMT CTN study 0802 (n = 213) [11], Hôpital Saint Louis, Paris (n = 184), and the University of Michigan (n = 307). Patients from Minnesota or Michigan enrolled in the BMT CTN trials (n = 33) were counted only once. The distribution of initial GVHD organ staging combinations and subsequent responses to therapy were analyzed to predict responses and define HR- and SR-GVHD.

PATIENTS AND METHODS

Study Design

Between 1990 and 2007, 1723 allogeneic HCT patients developed grade I-IV acute GVHD and were treated with prednisone 2 mg/kg/day or 60 mg/m^2 per oral (or methylprednisolone 48 mg/m^2 i.v.) as initial therapy and are included in this analysis. All HCT and data collection protocols were reviewed and approved by the Institutional Review Boards at the respective HCT centers. Deidentified data were compiled for analysis.

Patient and Transplant Characteristics

Patient demographics including year of transplant, recipient age, gender, cytomegalovirus (CMV) serostatus, and underlying diagnosis are shown in Table 1. Median patient age was 40 years (range, 0.2 - 76) and 24% were < 20 years of age. Standard risk disease was defined as acute leukemia in first or second complete remission, chronic myelogenous leukemia (CML) in first chronic phase, or myelodysplastic syndrome (MDS) without excess blasts. All other diseases were considered high risk.

Transplant characteristics including donor type and preparative therapy are shown in Table 1. Graft sources included HLA-identical sibling bone marrow (BM) or peripheral blood stem cells (PBSC; n = 598), HLA-mismatched related donor BM or PBSC, (n = 73), HLA-matched unrelated donor (URD) BM or PBSC (n = 626), HLA-mismatched URD BM or PBSC (n = 164), single or double umbilical cord blood (UCB) (n = 262). The majority (74%) of patients received myeloablative conditioning regimens.

GVHD Therapy and Measurement of Response to Prednisone

All patients received prednisone at approximately 2 mg/kg or 60 mg/m^2/day per oral (or methylprednisolone i.v. equivalent) as initial therapy, and 353 (20%) received additional agents for initial treatment. Patients enrolled in the BMT CTN 0302 trial received steroids and 1 of 4 additional agents (etanercept, mycophenolate mofetil, denileukin diftitox, or pentostatin) [10]. Those enrolled in the BMT CTN 0802 trial received steroids with mycophenolate mofetil or placebo [11].

Response was determined comparing the initial acute GVHD stage and grade in each organ to the best recorded stage and grade at day 28 (±7 days) after prednisone treatment was started. Complete response (CR) was defined as the complete resolution of acute GVHD manifestations in all organs, without need for secondary GVHD therapy. Partial response (PR) was defined as improvement in GVHD stage in all initially affected organs, without resolution in all organs, worsening in any other GVHD target organs, or need for secondary GVHD therapy. No response was defined as the same severity of GVHD in any organ or death, or the addition of secondary GVHD therapy before day 28. Patients who experienced a flare of acute GVHD before day 28 and required therapy with increased steroids or additional GVHD therapy were also considered to have no response. Progression was defined as worsening GVHD in at least 1 organ with or without improvement in any other organ.

Statistical Analysis

Simple proportions with 95% confidence intervals were used to estimate the rate of overall response defined as CR+PR or CR alone. Survival after GVHD treatment was analyzed by Kaplan-Meier estimates [12]. Non-relapse, treatment-related mortality (TRM) was analyzed using cumulative incidence estimates, treating relapse of the underlying HCT diagnosis as a competing risk [13]. Outcomes were evaluated with the log-rank test for trend or the simple log-rank test when comparing 2 categories.

Logistic regression was used to examine the independent association of factors with CR+PR and CR alone at day 28 after onset of GVHD. Cox regression was used to assess the independent association of the high- and standard-risk categories with the probability of 6-month overall survival [14], and Fine and Gray proportional hazards regression was used to assess the independent association of the categories with risks of TRM and chronic GVHD [15]. Clinical groupings, probabilities within organ staging categories, and recursive partitioning were used to identify the optimal cut-point among initial acute GVHD stage groupings for response at day 28 [16]. The net reclassification index (NRI), defined as the net proportion of events assigned to a higher risk category plus the net proportion of nonevents assigned to a lower risk category, was used to assess and quantify the net improvement in risk prediction by our refined acute GVHD risk determination as compared to our original GVHD risk score [17] or other GVHD grading schemes [2,9,18,19].

Table 1
Patient and Transplant Characteristics

Characteristic	n (%)
Total no. patients	1723
Year of transplant	
1990-1995	289 (17)
1996-2000	292 (17)
2001-2005	409 (24)
2006-2007	733 (43)
Age, yr	
<20	421 (24)
21-40	459 (27)
41+	843 (49)
Median (range)	40 (.2-76)
Gender	
Male	1067 (62)
Female	656 (38)
Disease	
Acute leukemia	741 (43)
CML	217 (13)
CLL/other leukemia	73 (4)
MDS/MPN	194 (11)
HL/NHL	236 (14)
Other malignancies	69 (4)
SAA/FA	81 (5)
Immune deficiency/hemoglobinopathy	28 (2)
Inborn error of metabolism	28 (2)
Disease risk*	
Standard	816 (47)
High	907 (53)
Donor type	
Matched Sibling BM/PBSC	598 (35)
Mismatched Related Donor BM/PBSC	73 (4)
Matched URD BM/PBSC	626 (36)
Mismatched URD BM/PBSC	164 (10)
Single or double UCB	262 (15)
Conditioning	
Myeloablative	1273 (74)
Reduced intensity	450 (26)

CML indicates chronic myelogenous leukemia; CLL, chronic lymphocytic leukemia; MDS, myelodysplastic syndrome; MPN, myeloproliferative neoplasm; HL, Hodgkin lymphoma; NHL, non-Hodgkin lymphoma; SAA, severe aplastic anemia; FA, Fanconi anemia; UCB, umbilical cord blood.

* Standard risk defined as acute leukemia in CR1 or CR2, CML in first chronic phase, MDS without excess blasts or non-malignant diseases. All others were high risk.

RESULTS

The median time from HCT to initiation of steroid therapy was 30 days (range, 2 to 178; interquartile range, 21 to 43). All patients had at least 3 months follow-up from GVHD treatment (median, 4.9 years; range .3 to 17.7).

The initial GVHD stage in each organ is shown in Table 2. Initial GVHD organ involvement was skin only (n = 910; 53%), upper and/or lower GI only (n = 346; 20%), liver only (n = 23; 1%), or multi-organ (n = 556; 25%). Overall response

Table 2
GVHD Organ Stage and Clinical Grade in 1723 Patients at Onset of Treatment

Organ Stage	0	1	2	3	4
Skin	400 (23%)	262 (15%)	409 (24%)	639 (37%)	13 (1%)
Liver	1579 (92%)	57 (3%)	52 (3%)	25 (1%)	10 (1%)
Lower GI	1197 (69%)	262 (15%)	122 (7%)	122 (7%)	20 (1%)
Upper GI	1263 (73%)	460 (27)			

	Initial CIBMTR Grade			
Initial MN Grade	A	B	C	D
I	146	280	0	0
II	0	383	570	0
III	0	121	183	7
IV	0	0	0	33

(CR+PR) at day 28 was observed in 1111 of 1723 patients (64%; 95% confidence interval [CI], 62% to 66%). CR was observed in 843 (49%, 95% CI, 47% to 51%) patients and PR in 268 (15%, 95% CI, 13% to 17%). For the entire cohort of 1723 patients, survival at 6 months after initiation of steroid therapy was 68% (95% CI, 66% to 70%) and TRM at 6 months was 26% (95% CI, 23% to 28%).

Initial GVHD grades by the Minnesota (MN) [2,9] and Center for International Blood and Marrow Transplant Research (CIBMTR) grading systems [19] at the initiation of steroid therapy are shown in Table 2. The day 28 CR/PR rate was 67% among patients with CIBMTR grades A and B GVHD and 61% among those with grades C and D GVHD ($P = .008$). The day 28 CR/PR rate was 68% among patients with MN grades I and II GVHD and 50% among those with grades III and IV GVHD ($P < .001$).

MN Acute GVHD Risk Score

According to a novel, previously published MN score that combined the MN and CIBMTR grading systems [9], 146 (8%) patients were classified as IA, 280 (16%) IB, 383 (22%) IIB, 570 (33%) IIC, 121 (7%) IIIB, 183 (11%) IIIC, 7 (<1%) IIID, and 33 (2%) IVD GVHD. This classification was used to categorize SR and HR groups. The day 28 CR/PR rate was 67% among 1500 patients with SR-GVHD (initial grade IA-IIIB) and 46% among 223 patients with HR-GVHD (initial grade IIIC-IVD) ($P <$ .001). In multivariate analysis, factors associated with CR/PR at day 28 after starting steroid treatment included GVHD risk group, donor type, and organ involvement, but not patient age. Additionally, factors associated with a higher likelihood of survival and lower TRM at 6 months included GVHD risk group, younger patient age, disease risk, donor type, and earlier onset of acute GVHD.

GVHD Organ Stage and Outcomes

Recognizing this heterogeneity in GVHD grading, particularly within clinical grades II and III, we then examined the details of initial GVHD organ stage combinations to determine whether stage groupings would better identify the patients at highest risk. We studied the 1723 patients and divided them into 67 categories by organ stage and, thus, extent of GVHD involvement at onset (Supplemental Table 1). We collapsed these categories into 17 larger categories clustered as clinically similar cohorts with comparable CR/PR at day 28, and we evaluated these new GVHD staging categories for CR/PR, survival, and TRM (Table 3). We found a clear demarcation between categories according to the CR/PR rate at day 28 and risk of 6-month mortality and TRM, thus dividing the cohort into SR-GVHD and HR-GVHD groups.

Refined Acute GVHD Risk Score

This refined definition of SR-GVHD includes single organ involvement (either stage 1-3 skin or stage 1-2 GI) or 2 organ involvement (either stage 1-3 skin plus stage 1 GI; or stage 1-3 skin plus stage 1-4 liver). All other patients are considered HR. Table 4 describes the stage and organ involvement (and subgroup frequency) by risk group for the patients in this analysis. The day 28 CR/PR rate was lower in the

Table 3
GVHD Organ Staging Categories

GVHD Categories Grouped by Day 28 CR/PR	n	Day 28 CR/PR	OR of Day 28 CR/PR (95% CI)	*P* Value	Relative Risk of Mortality (95% CI)	*P* Value	Relative Risk of TRM (95% CI)	*P* Value	GVHD Risk*
1. Stage 1-3 skin only†	901	68%	1.0		1.0		1.0		SR
2. UGI only	115	78%	1.6 (1.0-2.6)	.04	.6 (.4-.9)	.02	.5 (.3-.9)	.03	SR
3. Stage 1-2 lower GI only	100	73%	1.3 (.8-2.1)	.29	1.2 (.8-1.7)	.41	1.6 (1.0-2.3)	.04	SR
4. Stage 1-3 skin + UGI	90	69%	1.0 (.6-1.6)	.89	.9 (.6-1.4)	.64	.9 (.6-1.5)	.71	SR
5. Stage 1-3 skin + stage 1 Lower GI	71	61%	.7 (.4-1.2)	.16	1.1 (.7-1.8)	.60	1.3 (.8-2.1)	.35	SR
6. Stage 1 lower GI + upper GI	64	64%	.8 (.5-1.4)	.42	1.2 (.7-1.9)	.54	1.4 (.9-2.4)	.18	SR
7. Stage 1-3 skin + stage 1-4 liver	51	71%	1.1 (.6-2.1)	.71	1.3 (.8-2.0)	.30	1.4 (.9-2.3)	.17	SR
8. Stage 1-3 skin + stage 1 lower GI + upper GI	62	61%	.6 (.4-.9)	.12	.6 (.3-1.1)	.11	.6 (.3-1.3)	.18	SR
9. Stage 1-2 lower GI + stage 1-3 liver	12	50%	.4 (.1-1.4)	.15	2.9 (1.4-6.3)	.005	3.2 (1.5-7.0)	.003	HR
10. Stage 1-3 skin + (stage 1-2 lower GI or upper GI) + stage 1-3 liver	23	35%	.2 (.1-.6)	.001	3.0 (1.7-5.1)	<.001	3.2 (1.8-5.6)	<.001	HR
11. Stage 3 lower GI only	65	55%	.5 (.3-.9)	.02	2.1 (1.4-3.1)	.003	2.7 (1.8-4.2)	<.001	HR
12. Stage 1-3 skin + stage 2 lower GI	54	52%	.5 (.3-.9)	.01	1.5 (1.0-2.3)	.08	1.6 (1.0-2.6)	.06	HR
13. Stage 3-4 lower GI + (stage 1-3 skin or liver stage 1-4)	55	36%	.2 (.1-.4)	<.001	2.5 (1.7-3.6)	<.001	3.0 (2.0-4.5)	<.001	HR
14. Stage 1-4 liver alone	25	48%	.3 (.2-.8)	.01	1.0 (.5-2.3)	.96	1.7 (.8-3.8)	.19	HR
15. Stage 1-3 skin + stage 3-4 lower GI + stage 1-4 liver	13	8%	.1 (.01-.3)	.002	4.3 (2.3-8.2)	<.001	7.4 (3.6-15.2)	<.001	HR
16. Stage 4 skin only	13	38%	.3 (.1-.8)	.02	.8 (.2-3.1)	.71	2.2 (.7-6.9)	.16	HR
17. Stage 4 lower GI only	22	22%	.1 (.03-.7)	.02	3.5 (1.5-7.9)	.003	3.1 (1.2-8.1)	.02	HR

SR indicates standard risk; HR, high risk, UGI, upper gastrointestinal.

* SR and HR categories assigned by recursive partitioning for all 3 endpoints.

† Reference group; regression analysis includes adjustment for donor type [HLA-matched sibling versus other related versus umbilical cord blood versus HLA-matched URD versus HLA-mismatched URD], days to acute GVHD (<28 days versus ≥ 28 days), center/dataset (Minnesota versus BMTCTN 0302 versus BMTCTN 0802 versus Paris versus Michigan), age (<21 versus 21-40 versus 41+).

Table 4
GVHD Risk Definition by Organ Stage at Onset

GVHD Risk Score	One Organ (n)	Two Organs (n)	Three Organs (n)
SR (n = 1454, 84%)	Stage 1-3 Skin (901) Stage 1-2 GI (279)[†]	Stage 1-3 skin plus stage 1 GI (223)[*] Stage 1-3 skin plus stage 1-4 liver (51)	– –
HR[‡] (n = 269, 16%)	Stage 4 Skin (13) Stage 3-4 GI (74)[§] Stage 1-4 Liver (25)[‖]	Stage 1-3 skin plus stage 2 GI (54) Stage 1-2 lower GI plus stage 1-3 liver (12) Stage 3-4 GI plus stage 1-3 skin (45) Stage 3-4 GI plus stage 1-4 liver (10)	Stage 1-3 skin plus stage 1-2 GI plus stage 1-3 liver (23) Stage 1-3 skin plus stage 3-4 GI plus stage 1-4 liver (13)

UGI plus lower GI considered as single-organ disease.
[*] Stage 1-3 skin plus stage 1 GI includes stage 1-3 skin plus UGI (n = 90), stage 1-3 skin plus stage 1 lower GI (n = 71), or stage 1-3 skin plus UGI plus stage 1 lower GI (n = 62).
[†] Stage 1-2 GI includes UGI alone (n = 115), stage 1-2 lower GI alone (n = 100), or UGI or stage 1 lower GI (n = 64).
[‡] For HR disease, the degree of organ involvement is the minimum necessary to be deemed HR. Patients with higher stage of GVHD than observed in the HR group should also be considered HR.
[§] Stage 3-4 GI includes stage 3 lower GI alone (n = 65), or stage 4 lower GI alone (n = 9).
[‖] Stage 1-4 liver includes Stage 1 liver alone (n = 7), stage 2 liver alone (n = 10), stage 3 liver alone (n = 5), or stage 4 liver alone (n = 3).

269 patients with HR-GVHD than in the 1454 patients with SR-GVHD (44% [95% CI, 38% to 50%] versus 68% [95% CI, 66% to 70%], $P < .001$) (Figure 1). At day 28, the CR rates in HR-GVHD and SR-GVHD cohorts were 27% (95% CI, 22% to 33%) and 48% (95% CI, 45% to 51%); PR rates were 16% (95% CI, 12% to 21%) and 21% (95% CI, 19% to 23%). Additionally, the 6-month incidence of TRM was twice as high in patients with HR-GVHD than in those with SR-GVHD (44% [95% CI, 38% to 50%] versus 22% [95% CI, 20% to 24%], $P < .001$) (Figure 2). Finally, survival at 6 months after the onset of steroid treatment was lower in patients with HR-GVHD than in those with SR-GVHD (52% [95% CI, 46% to 58%] versus 71% [95% CI, 69% to 73%], $P < .001$).

In multiple regression analysis adjusted for clinically significant variables in addition to the risk group, the probability of CR/PR at day 28 in patients with HR-GVHD was lower than in those with SR-GVHD (odds ratio [OR], .3; 95% CI, .2 to .4; $P < .001$) (Table 5). Donor type was the only other factor associated with response. Patients who received a graft from an HLA-matched (OR, .7; 95% CI, .6 to .9; $P = .01$) or mismatched URD (OR, .3; 95% CI, .2 to .5; $P < .001$) were less likely to respond than those who received either a related donor or umbilical cord blood graft. The number of days from HCT to GVHD onset, patient age, original center/study cohort, and other factors showed no statistically significantly association with response.

In multivariate analysis, factors associated with CR/PR were also associated with 6-month survival and TRM. These included patient age, disease risk, donor type, time to onset of acute GVHD, and the new, refined GVHD risk group. Patients with HR-GVHD had a 2-fold increase in risk of mortality (relative risk, 2.1; 95% CI, 1.7 to 2.6; $P < .001$) and a 2.5-increased risk of TRM (relative risk, 2.5; 95% CI, 2.0 to 3.2; $P < .001$) compared with patients with SR-GVHD. Risks of mortality and TRM were also significantly higher in older patients, recipients of HLA-matched or mismatched URD grafts, and in those with onset of GVHD within 28 days after HCT (Table 5).

The refined GVHD risk score more effectively redistributes patients based on their predicted outcomes compared with the original GVHD risk score and the CIBMTR or the MN grading systems (Table 6). Seventy-six of 1500 (5%) SR patients in the original scoring system, 72 of 930 (8%) patients with CIBMTR grades A-B, and 14 (1%) of the 1379 patients with MN I-II GVHD were reclassified as HR according to this refined scoring system. More strikingly, 30 of 223 (13%) HR patients from the original scoring system, 596 of 793 (75%) patients with CIBMTR grades C-D, and 89 of 344 (26%) of MN grade III-IV were reclassified as SR according to this refined scoring system.

As shown in Table 7, the refined GVHD risk score better demarcates day 28 CR/PR (SR 68% versus HR 44%, $P < .001$) compared with our initial GVHD risk score (SR 67% versus HR 46%, $P < .001$), CIBMTR grading system (A-B 67% versus C-D 61%, $P = .008$), or the MN GVHD grading system (I-II 68% versus III-IV 50%, $P < .001$). Similarly, our new refined risk classification better predicts 6-month survival (SR 71%

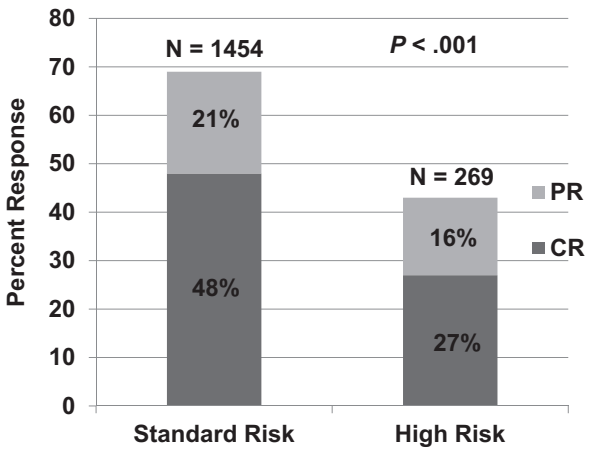

Figure 1. Day 28 response by risk group. CR, complete response; PR, partial response.

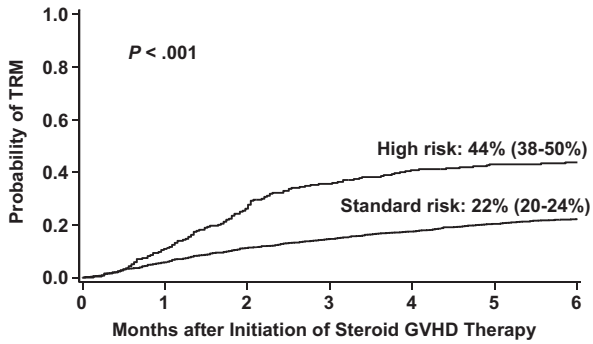

Figure 2. Cumulative incidence (±95% CI) of TRM at 6 months after onset of steroid therapy according to risk group.

Table 5
Factors Associated with Day 28 CR/PR, 6 Month Mortality and TRM: Multivariate Analysis

Factors	n	OR of Day 28 CR/PR (95% CI)	*P* Value	Relative Risk of Mortality (95% CI)	*P* Value	Relative Risk of TRM (95% CI)	*P* Value
Age							
<20*	421	1.0		1.0		1.0	
21-40†	459	.8 (.6-1.1)	.10	1.3 (1.0-1.8)	**.03**	1.6 (1.2-2.1)	**.004**
41+	843	.8 (.6-1.1)	.15	1.7 (1.3-2.3)	**<.001**	2.4 (1.8-3.3)	**<.001**
Disease risk‡							
Standard*	812	1.0		1.0		1.0	
High	717	1.1 (.9-1.4)	.45	1.2 (1.0-1.5)	**.05**	1.1 (.9-1.3)	.52
Nonmalignant	191	.9 (.6-1.3)	.47	1.06 (.8-1.5)	.74	1.8 (1.3-2.5)	**.001**
Conditioning regimen							
Myeloablative*	1273	1.0		1.0		1.0	
Reduced intensity	450	1.2 (.9-1.6)	.24	1.0 (.8-1.2)	.78	.9 (.6-1.1)	.25
Donor type							
Matched sibling*	598	1.0		1.0		1.0	
Other related	73	.9 (.5-1.5)	.62	1.5 (1.0-2.2)	.07	1.6 (1.0-2.6)	.42
Matched URD	626	.7 (.6-.9)	**.01**	1.5 (1.2-1.8)	**<.001**	1.6 (1.2-2.0)	**<.001**
Mismatched URD	164	.3 (.2-.5)	**<.001**	2.2 (1.6-2.9)	**<.001**	2.4 (1.7-3.2)	**<.001**
UCB	262	.9 (.7-1.3)	.45	1.1 (.8-1.4)	.67	1.0 (.7-1.4)	.81
Days from HCT to initial steroid treatment							
<28 days*	717	1.0		1.0		1.0	
≥28 days	1006	1.2 (1.0-1.5)	.06	.8 (.7-.9)	**.006**	.8 (.7-1.0)	**.02**
GVHD risk§							
SR*	1454	1.0		1.0		1.0	
HR	269	.3 (.2-.4)	**<.001**	2.1 (1.7-2.6)	**<.001**	2.5 (2.0-3.2)	**<.001**

Bold indicates statistical significance.

* Reference group.

† Comparison of 21-40 versus 41+: for CR/PR, $P = .27$; for mortality, $P = .02$; for TRM, $P < .001$.

‡ Standard-risk includes acute leukemia in CR1 or CR2, CML in first chronic phase, MDS without excess blasts, or non-malignant disease. High risk includes all others.

§ As defined in Table 3 with SR including organ involvement (stage 1-3 skin or stage 1-2 GI) or 2 organ involvement (stage 1-3 skin plus stage 1 GI; or stage 1-3 skin plus stage 1-4 liver). All other patients are HR.

versus HR 53%, $P < .001$) compared with our original GVHD risk score (SR 70% versus HR 54%, $P < .001$), the CIBMTR grading system (A-B 71% versus C-D 65%, $P = .008$) or the MN grading system (grade I-II 71% versus grade III-IV 56%, $P < .001$). TRM at 6 months was similar using this refined risk score (SR 22% versus HR 43%, $P < .001$) and the original GVHD risk score (SR 23% versus HR 42%, $P < .001$), but is better delineated than with the CIBMTR grading system (A-B 23% versus C-D: 29%, $P = .006$) or the MN grading system (grade I-II 22% versus grade III-IV 40%, $P < .001$). As measured by the NRI, our refined definition of GVHD improves both the true-positive and false-positive rates among our study population. The net percentage of patients more appropriately reclassified both in the positive and negative direction (eg, response and no response) is 5% ($P < .001$) for response and 4% for 6-month survival ($P < .001$) compared with our original definition of GVHD risk. Importantly, the NRI is 6% ($P < .001$) for response, 4% for 6-month survival ($P = .002$) and 5% for 6-month TRM ($P < .001$) compared to the MN grading system and is a better discriminant of outcome for all endpoints than the NRI of our original GVHD risk score and MN grading system.

DISCUSSION

In our previous analysis of GVHD scoring systems, we observed that the heterogeneity of the CIBMTR grades B-C (which include MN grades I-III) contributed to weak predictive utility for GVHD therapy outcomes [9]. We have, therefore, derived a combined scoring system, which better identified HR-GVHD patients who need more intensive initial therapy than prednisone alone [9]. To develop this refined GVHD risk score, we included independent data from 2 prospective BMT CTN trials as well as from 2 other large HCT centers (Hôpital Saint Louis, Paris and the University of Michigan). The refined definition of SR and HR groups, based upon initial GVHD stage, serves as a better predictor of response and survival than either the previous risk score, which was based upon initial GVHD grade [9], or the MN or CIBMTR grading systems. Patients in the HR group were much less likely to respond to initial steroid therapy and had at least 2-fold increased risk of mortality and TRM than those in the SR group. Outcomes are poor for patients with HR-GVHD treated with steroids alone, and these patients should be considered for enrollment in clinical trials testing new approaches for treatment of acute GVHD. Conversely,

Table 6
GVHD Risk Classification by Different Scoring Systems

Refined Acute GVHD Risk Score	Original Acute GVHD Risk Score [9]		CIBMTR Grading [20]		MN Grading [18,19]		n (%)
	SR	HR	SR (A-B)	HR (C-D)	SR (I-II)	HR (III-IV)	
SR	1424	30	858	596	1365	89	1454 (84%)
HR	76	193	72	197	14	255	269 (16%)
Total	1500 (87%)	223 (13%)	930 (54%)	793 (46%)	1379 (80%)	344 (20%)	

Table 7
Outcomes Based on Published GVHD Grading Systems

GVHD Grading System	n	CR/PR at Day 28	*P*	6-Month Survival	*P*	6-Month TRM	*P*
CIBMTR							
A-B	930	626 (67%)		660 (71%)		213 (23%)	
C-D	793	485 (61%)	.008	515 (65%)	.008	228 (29%)	.006
MN							
I-II	1379	938 (68%)		984 (71%)		304 (22%)	
III-IV	344	173 (50%)	<.001	192 (56%)	<.001	137 (40%)	<.001
Original GVHD risk score							
SR	1500	1008 (67%)		1055 (70%)		347 (23%)	
HR	223	103 (46%)	<.001	121 (54%)	<.001	94 (42%)	<.001
Refined GVHD risk score							
SR	1454	993 (68%)		1039 (71%)		326 (22%)	
HR	269	118 (44%)	<.001	137 (53%)	<.001	115 (43%)	<.001

	Number changed from SR- to HR-GVHD	Number changed from HR- to SR-GVHD	% Reclassified	*P*	% Reclassified	*P*	% Reclassified	*P*
NRI* (refined versus original)	76	30	5%	<.001	4%	<.001	0%	.21
NRI* (refined versus MN)	14	89	6%	<.001	4%	.002	5%	<.001
NRI* (refined versus CIBMTR)	72	596	19%	<.001	7%	.001	14%	<.001
NRI* (original versus MN)	0	121	1%	.74	2%	.02	4%	.001

* Net reclassification Index (NRI) defined as the net percentage of patients more appropriately reclassified both in positive and negative direction (eg, response and no response) [17]. The *P* value is comparing against the null hypothesis that there is no net benefit.

those with SR-GVHD should be considered for studies using therapeutic approaches designed to limit risks of treatment toxicity or risks of chronic GVHD.

Our refined definition of GVHD risk is confounded by a few small subsets of patients. Twenty-five (1%) patients with stage 1-4 liver disease alone were deemed HR-GVHD, as their outcomes were universally poor. However, 51 (3%) patients with stage 1-3 skin plus stage 1-4 liver had favorable outcomes and were scored SR-GVHD. Although this discrepancy cannot be fully explained, patients with liver only disease consistently had poor outcomes and should still be considered as having high risk disease while raising questions about the validity of their GVHD diagnoses compared with nonrecognized alternative hepatic syndromes complicating the diagnosis of liver only GVHD. Ten patients (<.5%) did not easily fit into any category, as shown in the Supplemental Table 1, and were included within a clinically similar category. Additionally, not all potential organ stage combinations and permutations were observed in our cohort of 1723 patients. Given the poor outcomes of patients in the HR-GVHD category, future patients presenting with higher stage involvement than those observed should also be considered high risk.

We chose 2 GVHD risk groups because when we examined CR/PR at day 28, as well as mortality and TRM, 2 groups became apparent: a large group consisting of 84% of the patients with similar outcomes and a smaller group comprising 16% of the patients with far worse outcomes. Within the large SR-GVHD group, there was no definable staging cohort with better outcomes. For instance the CR/PR rate of the 426 patients with stage 1-2 skin only GVHD (69%) was the same as the 475 patients with stage 3 skin only GVHD (67%). Therefore, although one might consider 3 groups to be more practical (ie, low, standard and high risk), the data showed that there were 2 GVHD risk groups, not 3 or more.

We also chose to use the entire cohort for analysis rather than divide into a training and validation set. Even in this

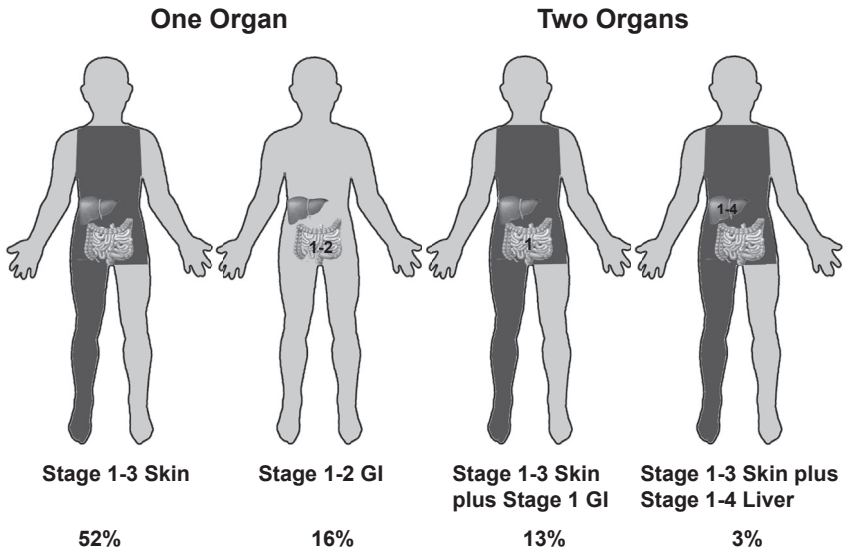

Figure 3. Organ stages defining HR acute GVHD with percentage involved of total cohort of 1723 patients.

largest data set ever analyzed for GVHD staging and grading characteristics, some of the identified groups are small. Subdivision of the population further would have either missed these cohorts entirely or limited our ability to classify them correctly. Future, independent analyses will thus better be able to confirm or refine the utility of this new classification.

It is simple to define SR-GVHD in our cohort (Table 4; Figure 3). These 84% patients were defined by single organ involvement (stage 1-3 skin or stage 1-2 GI) or 2 organ involvement (stage 1-3 skin plus stage 1 GI; or stage 1-3 skin plus stage 1-4 liver). Since our definition of HR-GVHD includes 9 distinct subgroups, we more simply define HR-GVHD as those who are not SR. These remaining 16% of patients had HR-GVHD characterized as having either severe single-organ involvement or complex multi-organ GVHD. We designed a free web-based program to easily determine the GVHD risk group for a given patient using our refined risk score, available at: http://z.umn.edu/MNAcuteGVHDRiskScore.

As more than 25,000 people worldwide develop acute GVHD annually, our new scoring system would reclassify 20,750 cases (83%) if the CIBMTR grading were used and 6750 (27%) if MN grading system were used, allowing for more appropriate therapy upfront. Biomarkers have been recently studied as a means to identify HR-GVHD patients, but their prognostic capability varies and no validated biomarker test is currently available for real-time decision making [6-8,21]. By contrast, this new revised clinical scoring system is available to clinicians at the bedside in real time. A future prospective study to examine this GVHD score along with informative biomarkers in the same GVHD population could help further refine the prognostic capability of each and determine whether a combination of clinical score and biomarker levels could even better identify HR-GVHD.

ACKNOWLEDGMENTS

The authors thank the nurses, nurse coordinators, and physicians who cared for these patients and their families. In addition, we gratefully acknowledge the research nurses at all sites whose dedicated efforts facilitated the prospective collection of the GVHD data. The authors also thank the patients and their families for participating in the clinical research trials.

Financial disclosure: This study was supported in part by the National Institutes of Health, National Cancer Institute grant 2P01CA065493 and the National Institutes of Health, Blood and Marrow Transplant Clinical Trials Network grant U10HL069290.

Conflict of interest statement: There are no conflicts of interest to report.

Authorship statement: M.L.M. and D.J.W. were responsible for the study concept and design. T.E.D. was responsible for the statistical design and analysis. M.L.M., M.R. A.C.H., A.A., V.T.H., J,B-M, J.L.M.F., R.J., M.A., B.R.B., S.G.H., D.J., M.P., G.S., J.H.A., J.E.L., and D.J.W were responsible for patient accrual and care. M.L.M. drafted the manuscript. All authors reviewed, edited and approved the manuscript.

SUPPLEMENTARY DATA

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.bbmt.2015.01.001

REFERENCES

1. Weisdorf D, Haake R, Blazar B, et al. Treatment of moderate/severe acute graft-versus-host disease after allogeneic bone marrow transplantation: an analysis of clinical risk features and outcome. *Blood*. 1990;75:1024-1030.
2. MacMillan ML, Weisdorf DJ, Wagner JE, et al. Response of 443 patients to steroids as primary therapy for acute graft-versus-host disease: comparison of grading systems. *Biol Blood Marrow Transplant*. 2002;8:387-394.
3. Martino R, Romero P, Subira M, et al. Comparison of the classic Glucksberg criteria and the IBMTR Severity Index for grading acute graft-versus-host disease following HLA-identical sibling stem cell transplantation. International Bone Marrow Transplant Registry. *Bone Marrow Transplant*. 1999;24:283-287.
4. Weisdorf DJ, Hurd D, Carter S, et al. Prospective grading of graft-versus-host disease after unrelated donor marrow transplantation: a grading algorithm versus blinded expert panel review. *Biol Blood Marrow Transplant*. 2003;9:512-518.
5. Cahn JY, Klein JP, Lee SJ, et al. Prospective evaluation of 2 acute graft-versus-host (GVHD) grading systems: a joint Societe Francaise de Greffe de Moelle et Therapie Cellulaire (SFGM-TC), Dana Farber Cancer Institute (DFCI), and International Bone Marrow Transplant Registry (IBMTR) prospective study. *Blood*. 2005;106:1495-1500.
6. Levine JE, Logan BR, Wu J, et al. Acute graft-versus-host disease biomarkers measured during therapy can predict treatment outcomes: a Blood and Marrow Transplant Clinical Trials Network study. *Blood*. 2012;119:3854-3860.
7. Vander Lugt MT, Braun TM, Hanash S, et al. ST2 as a marker for risk of therapy-resistant graft-versus-host disease and death. *N Engl J Med*. 2013;369:529-539.
8. Levine JE, Harris AC, Taylor A, et al. A biomarker-based grading system at onset of GVHD predicts TRM better than the modified Glucksberg Grading System. *Blood*. 2013;122:145.
9. MacMillan ML, Defor TE, Weisdorf DJ. What predicts high risk acute graft-versus-host disease (GVHD) at onset?: identification of those at highest risk by a novel acute GVHD risk score. *Br J Haematol*. 2012;157:732-741.
10. Alousi AM, Weisdorf DJ, Logan BR, et al. Etanercept, mycophenolate, denileukin, or pentostatin plus corticosteroids for acute graft-versus-host disease: a randomized phase 2 trial from the Blood and Marrow Transplant Clinical Trials Network. *Blood*. 2009;114:511-517.
11. Bolaños-Meade J, Logan BR, Alousi AM, et al. Phase 3 clinical trial of steroids/mycophenolate mofetil vs steroids/placebo as therapy for acute GVHD: BMT CTN 0802. *Blood*. 2014;124(22):3221-3227.
12. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53:457-481.
13. Lin DY. Non-parametric inference for cumulative incidence functions in competing risks studies. *Stat Med*. 1997;16:901-910.
14. Cox DR. Regression models and life tables. *J Royal Stast Soc*. 1972;34:187-220.
15. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94:496-509.
16. Breiman L, Freidman J, Olshen R, Stone C. *Classification and regression trees*. Belmont, CA: Wadsworth and Brooks; 1984.
17. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008;27:157-172. discussion 207-212.
18. Przepiorka D, Weisdorf D, Martin P, et al. 1994 Consensus Conference on Acute GVHD Grading. *Bone Marrow Transplant*. 1995;15:825-828.
19. Rowlings PA, Przepiorka D, Klein JP, et al. IBMTR Severity Index for grading acute graft-versus-host disease: retrospective comparison with Glucksberg grade. *Br J Haematol*. 1997;97:855-864.
20. Weisdorf DJ, Snover DC, Haake R, et al. Acute upper gastrointestinal graft-versus-host disease: clinical significance and response to immunosuppressive therapy. *Blood*. 1990;76:624-629.
21. Paczesny S. Discovery and validation of graft-versus-host disease biomarkers. *Blood*. 2013;121:585-594.